PHOTOCLEAVABLE FLUORESCENT NUCLEOTIDES FOR DNA SEQUENCING ON CHIP CONSTRUCTED BY SITE-SPECIFIC COUPLING CHEMISTRY

This application claims the benefit of copending U.S. Provisional Application No. 60/550,007, filed March 3, 2004, the contents of which are hereby incorporated by reference.

The invention disclosed herein was made with Government support under Center of Excellence in Genomic Science grant number P50 HG002806 from the National Institutes of Health, U.S. Department of Health and Human Services. Accordingly, the U.S. Government has certain rights in this invention.

Throughout this application, various publications are referenced in parentheses by number. Full citations for these references may be found at the end of each experimental section. The disclosures of these publications in their entireties are hereby incorporated by reference into this application to more fully describe the state of the art to which this invention pertains.

Background

DNA sequencing is a fundamental tool for biological science. The completion of the Human Genome Project has set the stage for screening genetic mutations to identify disease genes on a genome-wide scale (1). Accurate high-throughput DNA sequencing methods are needed to explore

the complete human genome sequence for applications in clinical medicine and health care. Recent studies have indicated that an important route for identifying functional elements in the human genome involves sequencing the genomes of many species representing a wide sampling of the evolutionary tree (2). To overcome the limitations of the current electrophoresis-based sequencing technology (3-5), a variety of new DNA-sequencing methods have been investigated. Such approaches include sequencing by hybridization (6), mass spectrometry based sequencing (7-9), and sequence-specific detection of single-stranded DNA using engineered nanopores (10). More recently, DNA sequencing by synthesis (SBS) approaches such as pyrosequencing (11), sequencing of single DNA molecules (12) and polymerase colonies (13) have been widely explored.

The concept of DNA sequencing by synthesis was revealed in 1988 (14). This approach involves detection of the identity of each nucleotide immediately after its incorporation into a growing strand of DNA in a polymerase reaction. Thus far, no complete success has been reported in using such a system to sequence DNA unambiguously. An SBS approach using photocleavable fluorescent nucleotide analogues on a surface was proposed in 2000 (15). In this approach, modified nucleotides are used as reversible terminators, in which a different fluorophore with a distinct fluorescent emission is linked to each of the 4 bases through a photocleavable linker and the 3'-OH group is capped by a small chemical moiety. DNA polymerase incorporates only a single nucleotide analogue complementary to the base on a

DNA template covalently linked to a surface. After incorporation, the unique fluorescence emission is detected to identify the incorporated nucleotide and the fluorophore is subsequently removed photochemically. The 3'-OH group is then chemically regenerated, which allows the next cycle of the polymerase reaction to proceed. Since the large surface on a DNA chip can have a high density of different DNA templates spotted, each cycle can identify many bases in parallel, allowing the simultaneous sequencing of a large number of DNA molecules. The advantage of using photons as reagents for initiating photoreactions to cleave the fluorophore is that no additional chemical reagents are required to be introduced into the system and clean products can be generated with no need for subsequent purification.

Summary

This invention provides a method for determining the sequence of a DNA, wherein (i) about 1000 or fewer copies of the DNA are bound to a solid substrate via 1,3-dipolar azide-alkyne cycloaddition chemistry and (ii) each copy of the DNA comprises a self-priming moiety, comprising performing the following steps for each nucleic acid residue of the DNA to be sequenced:

(a) contacting the bound DNA with DNA polymerase and four photocleavable fluorescent nucleotide analogues under conditions permitting the DNA polymerase to catalyze DNA synthesis, wherein (i) the nucleotide analogues consist of an analogue of G, an analogue of C, an analogue of T and an analogue of A, so that a nucleotide analogue complementary to the residue being sequenced is bound to the DNA by the DNA polymerase, and (ii) each of the four analogues has a pre-determined fluorescence wavelength which is different than the fluorescence wavelengths of the other three analogues;

(b) removing unbound nucleotide analogues; and

(c) determining the identity of the bound nucleotide analogue,

thereby determining the sequence of the DNA.

This invention also provides a method for determining the sequence of an RNA, wherein (i) about 1000 or fewer copies of the RNA are bound to a solid substrate via 1,3-dipolar azide-alkyne cycloaddition chemistry and (ii)

each copy of the RNA comprises a self-priming moiety, comprising performing the following steps for each nucleic acid residue of the RNA to be sequenced:

(a) contacting the bound RNA with RNA polymerase and four photocleavable fluorescent nucleotide analogues under conditions permitting the RNA polymerase to catalyze RNA synthesis, wherein (i) the nucleotide analogues consist of an analogue of G, an analogue of C, an analogue of U and an analogue of A, so that a nucleotide analogue complementary to the residue being sequenced is bound to the RNA by the RNA polymerase, and (ii) each of the four analogues has a pre-determined fluorescence wavelength which is different than the fluorescence wavelengths of the other three analogues;

(b) removing unbound nucleotide analogues; and

(c) determining the identity of the bound nucleotide analogue,

thereby determining the sequence of the RNA.

This invention also provides a composition of matter comprising a solid substrate having a DNA bound thereto via 1,3-dipolar azide-alkyne cycloaddition chemistry, wherein (i) about 1000 or fewer copies of the DNA are bound to the solid substrate, and (ii) each copy of the DNA comprises a self-priming moiety.

This invention also provides a composition of matter comprising a solid substrate having an RNA bound thereto via 1,3-dipolar azide-alkyne cycloaddition chemistry, wherein (i) about 1000 or fewer copies of the RNA are

bound to the solid substrate, and (ii) each copy of the RNA comprises a self-priming moiety.

Brief Description of the Figures

Figure 1: DNA extension reaction performed in solution phase to characterize the 4 different photocleavable fluorescent nucleotide analogues (dUTP-PC-R6G, dGTP-PC-Bodipy-FL-510, dATP-PC-ROX, dCTP-PC-Bodipy-650). After each extension reaction, the DNA extension product (SEQ ID NOs. 1-4) is purified by HPLC for MALDI-TOF MS measurement to verify that it is the correct extension product. Photolysis is performed to produce a DNA product that is used as a primer for the next DNA extension reaction.

Figure 2. Polymerase extension scheme. Primer extended with dUTP-PC-R6G (1), and its photocleavage product 2; Product 2 extended with dGTP-PC-Bodipy-FL-510 (3), and its photocleavage product 4; Product 4 extended with dATP-PC-ROX (5), and its photocleavage product 6; Product 6 extended with dCTP-PC-Bodipy-650 (7), and its photocleavage product 8. After 10 seconds of irradiation with a laser at 355 nm, photocleavage is complete with all the fluorophores cleaved from the extended DNA products.

Figure 3. Panels (1)-(8). MALDI-TOF MS spectra of the four consecutive extension products and their photocleavage products. Primer extended with dUTP-PC-R6G (1), and its photocleavage product 2; Product 2 extended with dGTP-PC-Bodipy-FL-510 (3), and its photocleavage product 4; Product 4 extended with dATP-PC-ROX (5), and its photocleavage product 6; Product 6 extended with

dCTP-PC-Bodipy-650 (7), and its photocleavage product 8. After 10 seconds of irradiation with a laser at 355 nm, photocleavage is complete with all the fluorophores cleaved from the extended DNA products.

Figure 4. Immobilization of an azido-labeled PCR product on an alkynyl-functionalized surface and a ligation reaction between the immobilized single-stranded DNA template and a loop primer to form a self-priming DNA moiety on the chip. The sequence of the loop primer is shown in (A).

Figure 5. Schematic representation of SBS on a chip using four PC fluorescent nucleotides (Upper panel) and the scanned fluorescence images for each step of SBS on a chip (Lower panel). (1) Incorporation of dATP-PC-ROX; (2) Photocleavage of PC-ROX; (3) Incorporation of dGTP-PC-Bodipy-FL-510; (4) Photocleavage of PC-Bodipy-FL-510; (5) Incorporation of dATP-PC-ROX; (6) Photocleavage of PC-ROX; (7) Incorporation of dCTP-PC-Bodipy-650; (8) Photocleavage of PC-Bodipy-650; (9) Incorporation of dUTP-PC-R6G; (10) Photocleavage of PC-R6G; (11) Incorporation of dATP-PC-ROX; (12) Photocleavage of PC-ROX; (13) Incorporation of dUTP-PC-R6G; (14) Photocleavage of PC-R6G; (15) Incorporation of dATP-PC-ROX; (16) Photocleavage of PC-ROX; (17) Incorporation of dGTP-PC-Bodipy-FL-510; (18) Photocleavage of PC-Bodipy-FL-510; (19) Incorporation of dUTP-PC-R6G; (20) Photocleavage of PC-R6G; (21) Incorporation of dCTP-PC-Bodipy-650; (22) Photocleavage of PC-Bodipy-650; (23) Incorporation of dATP-PC-ROX; (24) Photocleavage of PC-ROX.

Figure 6. Structures of dGTP-PC-Bodipy-FL-510 ($\lambda_{abs\ (max)}$ = 502 nm; $\lambda_{em\ (max)}$ = 510 nm), dUTP-PC-R6G ($\lambda_{abs\ (max)}$ = 525 nm; $\lambda_{em\ (max)}$ = 550 nm), dATP-PC-ROX ($\lambda_{abs\ (max)}$ = 575 nm; $\lambda_{em\ (max)}$ = 602 nm), and dCTP-PC-Bodipy-650 ($\lambda_{abs\ (max)}$ = 630 nm; $\lambda_{em\ (max)}$ = 650 nm).

Figure 7. Synthesis of photocleavable fluorescent nucleotides. (a) acetonitrile or DMF/1 M $NaHCO_3$ solution; (b) *N,N'*-disuccinimidyl carbonate (DSC), triethylamine; (c) 0.1 M $Na_2CO_3/NaHCO_3$ aqueous buffer (pH 8.5-8.7).

Detailed Description of the Invention

Terms

The following definitions are presented as an aid in understanding this invention:

A	-	Adenine;
C	-	Cytosine;
DNA	-	Deoxyribonucleic acid;
G	-	Guanine;
RNA	-	Ribonucleic acid;
SBS	-	Sequencing by synthesis;
T	-	Thymine; and
U	-	Uracil.

"Nucleic acid " shall mean any nucleic acid, including, without limitation, DNA, RNA and hybrids thereof. The nucleic acid bases that form nucleic acid molecules can be the bases A, C, G, T and U, as well as derivatives thereof. Derivatives of these bases are well known in the art, and are exemplified in PCR Systems, Reagents and Consumables (Perkin Elmer Catalogue 1996 1997, Roche Molecular Systems, Inc., Branchburg, New Jersey, USA).

As used herein, "self-priming moiety" shall mean a nucleic acid moiety covalently bound to a nucleic acid to be transcribed, wherein the bound nucleic acid moiety, through its proximity with the transcription initiation site of the nucleic acid to be transcribed, permits transcription of the nucleic acid under nucleic acid polymerization-permitting conditions (e.g. the presence

of a suitable polymerase, nucleotides and other reagents). That is, the self-priming moiety permits the same result (i.e. transcription) as does a non-bound primer. In one embodiment, the self-priming moiety is a single stranded nucleic acid having a hairpin structure. Examples of such self-priming moieties are shown in the Figures.

"Hybridize" shall mean the annealing of one single-stranded nucleic acid to another nucleic acid based on sequence complementarity. The propensity for hybridization between nucleic acids depends on the temperature and ionic strength of their milieu, the length of the nucleic acids and the degree of complementarity. The effect of these parameters on hybridization is well known in the art (see Sambrook J, Fritsch EF, Maniatis T. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, New York.).

As used herein, "nucleotide analogue" shall mean an analogue of A, G, C, T or U which is recognized by DNA or RNA polymerase (whichever is applicable) and incorporated into a strand of DNA or RNA (whichever is appropriate). Examples of nucleotide analogues include, without limitation 7-deaza-adenine, 7-deaza-guanine, the analogues of deoxynucleotides shown in Figure 6, analogues in which a label is attached through a cleavable linker to the 5-position of cytosine or thymine or to the 7-position of deaza-adenine or deaza-guanine, analogues in which a small chemical moiety such as $-CH_2OCH_3$ or $-CH_2CH=CH_2$ is used to cap the -OH group at the

3'-position of deoxyribose, and analogues of related dideoxynucleotides. Nucleotide analogues, including dideoxynucleotide analogues, and DNA polymerase-based DNA sequencing are also described in U.S. Patent No. 6,664,079.

Embodiments of the Invention

This invention provides a method for determining the sequence of a DNA, wherein (i) about 1000 or fewer copies of the DNA are bound to a solid substrate via 1,3-dipolar azide-alkyne cycloaddition chemistry and (ii) each copy of the DNA comprises a self-priming moiety, comprising performing the following steps for each nucleic acid residue of the DNA to be sequenced:

(a) contacting the bound DNA with DNA polymerase and four photocleavable fluorescent nucleotide analogues under conditions permitting the DNA polymerase to catalyze DNA synthesis, wherein (i) the nucleotide analogues consist of an analogue of G, an analogue of C, an analogue of T and an analogue of A, so that a nucleotide analogue complementary to the residue being sequenced is bound to the DNA by the DNA polymerase, and (ii) each of the four analogues has a pre-determined fluorescence wavelength which is different than the fluorescence wavelengths of the other three analogues;

(b) removing unbound nucleotide analogues; and

(c) determining the identity of the bound nucleotide analogue,

thereby determining the sequence of the DNA.

In one embodiment, the instant method further comprises the step of photocleaving the fluorescent moiety from the bound nucleotide analogue following step (c).

In another embodiment of the instant method, the solid substrate is glass or quartz.

In a further embodiment of the instant method, fewer than 100 copies of the DNA, fewer than 20 copies of the DNA, or fewer than five copies of the DNA are bound to the solid substrate.

In still a further embodiment, one copy of the DNA is bound to the solid substrate.

This invention also provides a method for determining the sequence of an RNA, wherein (i) about 1000 or fewer copies of the RNA are bound to a solid substrate via 1,3-dipolar azide-alkyne cycloaddition chemistry and (ii) each copy of the RNA comprises a self-priming moiety, comprising performing the following steps for each nucleic acid residue of the RNA to be sequenced:

(a) contacting the bound RNA with RNA polymerase and four photocleavable fluorescent nucleotide analogues under conditions permitting the RNA polymerase to catalyze RNA synthesis, wherein (i) the nucleotide analogues consist of an analogue of G, an analogue of C, an analogue of U and an analogue of A, so that a nucleotide analogue complementary to the residue being sequenced is bound to the RNA by the RNA polymerase, and (ii)

each of the four analogues has a pre-determined fluorescence wavelength which is different than the fluorescence wavelengths of the other three analogues;

(b) removing unbound nucleotide analogues; and

(c) determining the identity of the bound nucleotide analogue,

thereby determining the sequence of the RNA.

In one embodiment the instant method, further comprises the step of photocleaving the fluorescent moiety from the bound nucleotide analogue following step (c).

In another embodiment of the instant method, the solid substrate is glass or quartz.

In a further embodiment of the instant method, fewer than 100 copies of the RNA, fewer than 20 copies of the RNA, or fewer than five copies of the RNA are bound to the solid substrate.

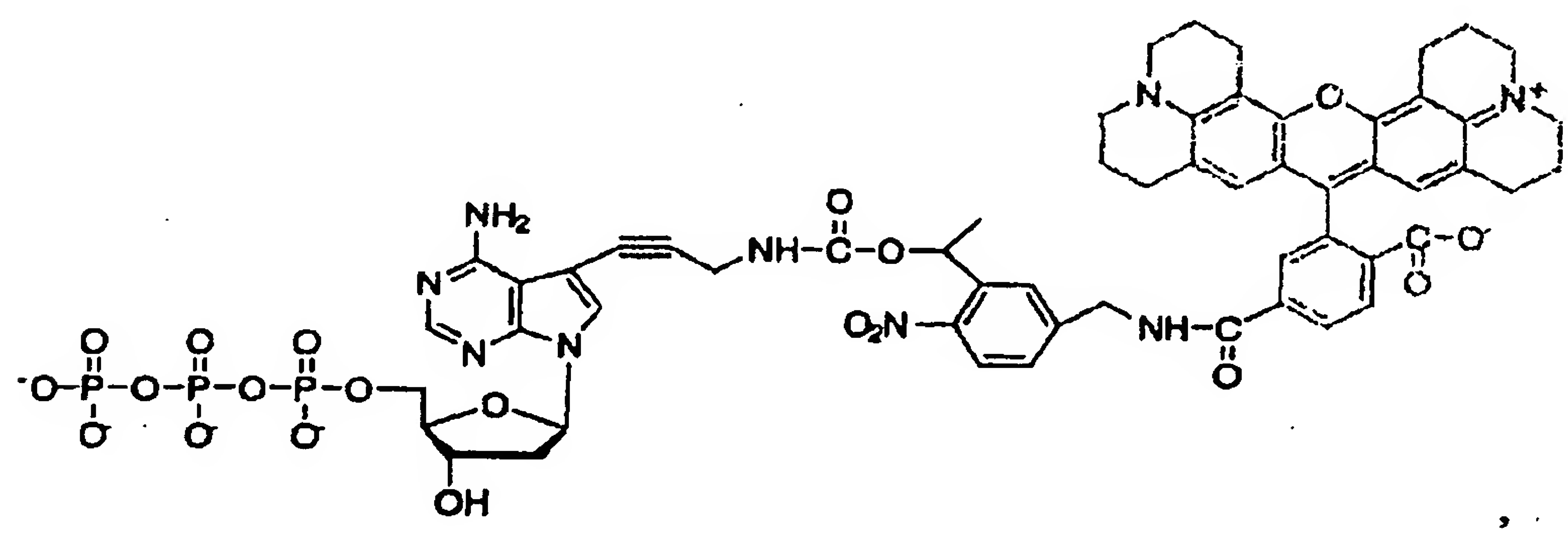
In still a further embodiment, one copy of the RNA is bound to the solid substrate.

This invention also provides a composition of matter comprising a solid substrate having a DNA bound thereto via 1,3-dipolar azide-alkyne cycloaddition chemistry, wherein (i) about 1000 or fewer copies of the DNA are bound to the solid substrate, and (ii) each copy of the DNA comprises a self-priming moiety.

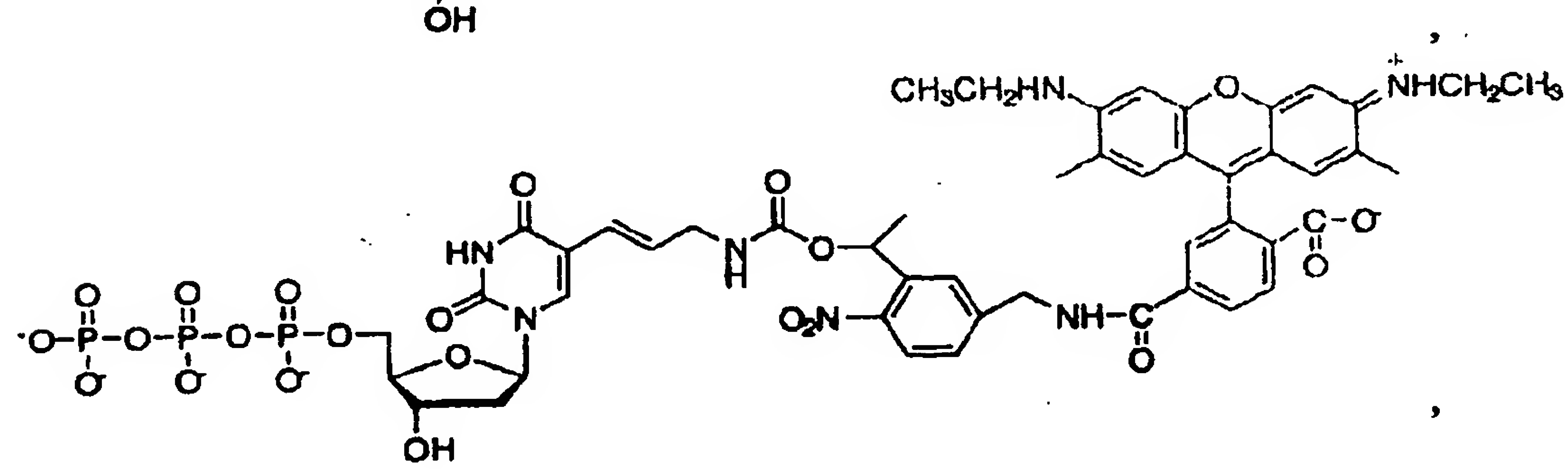
This invention also provides a composition of matter comprising a solid substrate having an RNA bound thereto via 1,3-dipolar azide-alkyne cycloaddition chemistry, wherein (i) about 1000 or fewer copies of the RNA are bound to the solid substrate, and (ii) each copy of the RNA comprises a self-priming moiety.

In one embodiment of the instant methods and compositions of matter, the number of DNA or RNA copies bound to the solid substrate exceeds 1000, and can be, for example, about 10^4, 10^5, 10^6, 10^7, 10^8, 10^9, or greater.

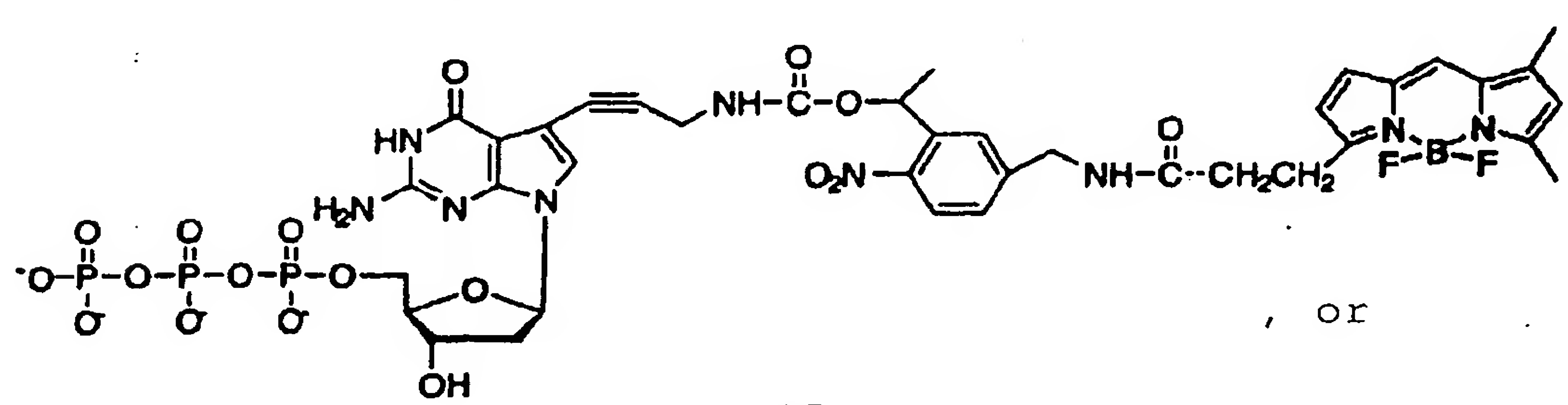
This invention also provides a compound having the structure:



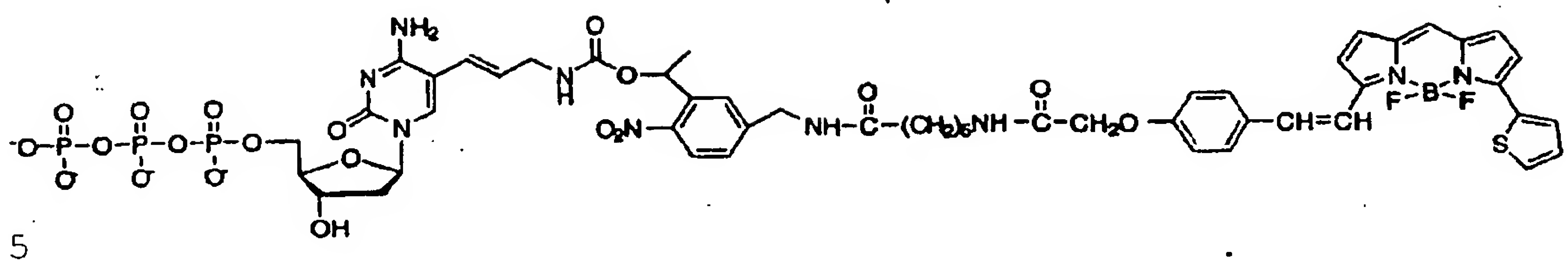
,



,



, or



This invention will be better understood by reference to the Experimental Details which follow, but those skilled in the art will readily appreciate that the specific experiments detailed are only illustrative of the invention as described more fully in the claims which follow thereafter.

Experimental Details

Synopsis

Here, the procedure for performing SBS on a chip using a synthetic DNA template and photocleavable pyrimidine nucleotides (C and U) is disclosed (also see 16). In addition, the design and synthesis of 4 photocleavable nucleotide analogues (A, C, G, U), each of which contains a unique fluorophore with a distinct fluorescence emission is described. Initially, it was established that these nucleotides are good substrates for DNA polymerase in a solution-phase DNA extension reaction and that the fluorophore can be removed with high speed and efficiency by laser irradiation (λ~355 nm). SBS was then performed using these 4 photocleavable nucleotide analogues to identify the sequence of a DNA template immobilized on a chip. The DNA template was produced by PCR using an azido-labeled primer, and was immobilized on the surface of the chip with 1,3-dipolar azide-alkyne cycloaddition chemistry. A self-priming moiety was then covalently attached to the DNA template by enzymatic ligation to allow the polymerase reaction to proceed on the DNA immobilized on the surface. This is the first report of using a complete set of photocleavable fluorescent nucleotides for 4-color DNA sequencing by synthesis.

Introduction

A 4-color DNA sequencing by synthesis (SBS) on a chip using four photocleavable fluorescent nucleotide

analogues (dGTP-PC-Bodipy-FL-510, dUTP-PC-R6G, dATP-PC-ROX, and dCTP-PC-Bodipy-650) is disclosed herein. Each nucleotide analogue consists of a different fluorophore attached to the 5-position of the pyrimidines (C and U) and the 7-position of the purines (G and A) through a photocleavable 2-nitrobenzyl linker. After verifying that these nucleotides could be successfully incorporated into a growing DNA strand in a solution-phase polymerase reaction and the fluorophore could be cleaved using laser irradiation (λ~355 nm) in 10 seconds, an SBS reaction was performed on a chip which contained a self-priming DNA template covalently immobilized using 1,3-dipolar azide-alkyne cycloaddition. The DNA template was produced by a polymerase chain reaction using an azido-labeled primer and the self-priming moiety was attached to the immobilized DNA template by enzymatic ligation. Each cycle of SBS consisted of the incorporation of the photocleavable fluorescent nucleotide into the DNA, detection of the fluorescent signal and photocleavage of the fluorophore. The entire process was repeated to identify 12 continuous bases in the DNA template. These results demonstrate that photocleavable fluorescent nucleotide analogues can be incorporated accurately into a growing DNA strand during a polymerase reaction in solution phase as well as on a chip. Moreover, all 4 fluorophores can be detected and then efficiently cleaved using near-UV irradiation, thereby allowing continuous identification of the DNA template sequence.

To demonstrate the feasibility of carrying out DNA sequencing by synthesis on a chip, four photocleavable fluorescent nucleotide analogues (dGTP-PC-Bodipy-FL-510,

dUTP-PC-R6G, dATP-PC-ROX, and dCTP-PC-Bodipy-650) (Fig. 6) were synthesized according to the scheme shown in Fig. 7 using a similar procedure as reported previously (16). Modified DNA polymerases have been shown to be highly tolerant to nucleotide modifications with bulky groups at the 5-position of pyrimidines (C and U) and the 7-position of purines (A and G) (17, 18). Thus, each unique fluorophore was attached to the 5 position of C/U and the 7 position of A/G through a photocleavable 2-nitrobenzyl linker.

In order to verify that these fluorescent nucleotides are incorporated accurately in a base-specific manner in a polymerase reaction, four continuous steps of DNA extension and photocleavage by near UV irradiation were carried out in solution as shown in Figure 1. This allows the isolation of the DNA product at each step for detailed molecular structure characterization as shown in Fig. 2. The first extension product 5'-U(PC-R6G)-3' **1** was purified by HPLC and analyzed using MALDI-TOF MS [Fig. 3]. This product was then irradiated at 355 nm using an Nd-YAG laser for 10 seconds and the photocleavage product **2** was also analyzed using MALDI-TOF MS [Fig. 3]. Near UV light absorption by the aromatic 2-nitrobenzyl linker causes reduction of the 2-nitro group to a nitroso group and an oxygen insertion into the carbon-hydrogen bond followed by cleavage and decarboxylation (19). As can be seen from Fig. 3, Panel 1, the MALDI-TOF MS spectrum consists of a distinct peak at *m/z* 6536 corresponding to the DNA extension product 5'--U(PC-R6G)-3' (**1**), which confirms that the nucleotide analogue can be incorporated base specifically by DNA polymerase into a growing DNA

strand. The small peak at m/z 5872 corresponding to the photocleavage product is due to the partial cleavage caused by the nitrogen laser pulse (337 nm) used in MALDI ionization. For photocleavage, a Nd-YAG laser was used to irradiate the DNA product carrying the fluorescent nucleotide for 10 seconds at 355 nm to cleave the fluorophore from the DNA extension product. Fig.3, Panel 2, shows the photocleavage result of the above DNA product. The peak at m/z 6536 has completely disappeared while the peak corresponding to the photocleavage product 5'-U (**2**) appears as the sole dominant peak at m/z 5872, which establishes that laser irradiation completely cleaves the fluorophore with high speed and efficiency without damaging the DNA. The next extension reaction was carried out using this photocleaved DNA product as a primer along with dGTP-PC-Bodipy-FL-510 to yield an extension product 5'-UG(PC-Bodipy-FL-510)-3' (**3**). As described above, the extension product **3** was purified, analyzed by MALDI-TOF MS producing a dominant peak at m/z 6751 [Fig. 3, Panel 3], and then photocleaved for further MS analysis yielding a single peak at m/z 6255 (product **4**) [Fig. 3, Panel 4]. The third extension using dATP-PC-ROX to yield 5'-UGA(PC-ROX)-3' (**5**), the fourth extension using dCTP-PC-Bodipy-650 to yield 5'-UGAC(PC-Bodipy-650)-3' (**7**) and their photocleavage to yield products **6** and **8** were similarly carried out and analyzed by MALDI-TOF MS as shown in Fig. 3, Panels 5-8. These results demonstrate that the above-synthesized four photocleavable fluorescent nucleotide analogues can successfully incorporate into the growing DNA strand in a polymerase reaction, and the fluorophore can be

efficiently cleaved by near UV irradiation, which makes it feasible to use them for SBS on a chip.

The photocleavable fluorescent nucleotide analogues were then used in an SBS reaction to identify the sequence of the DNA template immobilized on a solid surface as shown in Fig. 4. A site-specific 1,3-dipolar cycloaddition coupling chemistry was used to covalently immobilize the azido-labeled double-stranded PCR products on the alkynyl-functionalized surface in the presence of a Cu(I) catalyst. Previously, we have shown that DNA is successfully immobilized on the glass surface by this chemistry and evaluated the functionality of the surface-bound DNA and the stability of the array using a primer extension reaction (16). The surface-immobilized double stranded PCR product was denatured using a 0.1 M NaOH solution to remove the complementary strand without the azido group, thereby generating a single-stranded PCR template on the surface. Then, a 5′-phosphorylated self-priming moiety (loop primer) was ligated to the 3′-end of the above single stranded DNA template using *Taq* DNA ligase (20). The structure of the loop primer was designed to bear a thermally stable loop (21) and stem sequence with a melting temperature of 89 °C. The 12-bp overhanging portion of the loop primer was made complementary to the 12-bp sequence of the template at its 3′ end to allow the *Taq* DNA ligase to seal the nick between the 5′-phosphate group of the loop primer and the 3′-hydroxyl group of the single-stranded DNA template. This produces a unique DNA moiety that can self-prime for the synthesis of a complementary strand. The ligation was found to be in quantitative yield in a parallel solution-

phase reaction using the same primer and single-stranded DNA template.

The principal advantage offered by the use of a self-priming moiety as compared to using separate primers and templates is that the covalent linkage of the primer to the template in the self-priming moiety prevents any possible dissociation of the primer from the template under vigorous washing conditions. Furthermore, the possibility of mis-priming is considerably reduced and a universal loop primer can be used for all the templates allowing enhanced accuracy and ease of operation. SBS was performed on the chip-immobilized DNA template using the 4 photocleavable fluorescent nucleotide analogues, see Fig. 5. The structure of the self-priming DNA moiety is shown schematically in the upper panel, with the first 12 nucleotide sequence immediately after the priming site. The sequencing reaction on the chip was initiated by extending the self-priming DNA using dATP-PC-ROX (complementary to the T on the template), and Thermo Sequenase DNA polymerase. After washing, the extension of the primer by a single fluorescent nucleotide was confirmed by observing an orange signal (the emission signal from ROX) in a microarray scanner [Fig. 5, (1)]. After detection of the fluorescent signal, the surface was irradiated at 355 nm for 1 min using an Nd-YAG laser to cleave the fluorophore. The surface was then washed, and a negligible residual fluorescent signal was detected to confirm complete photocleavage of the fluorophore [Fig. 5, (2)]. This was followed by incorporation of the next fluorescent nucleotide complementary to the subsequent base on the template. The entire process of

incorporation, detection and photocleavage was performed multiple times using the four photocleavable fluorescent nucleotide analogues to identify 12 successive bases in the DNA template. The integrated fluorescence intensity on the spot, obtained from the scanner software, indicated that the incorporation efficiency was over 90% and more than 97% of the original fluorescence signal was removed by photocleavage. A negative control experiment consisting of incubating the self-priming DNA moiety with dATP-PC-ROX in the absence of DNA polymerase and washing the surface showed that negligible fluorescence remained as compared to that of Fig. 5, (1).

In summary, four photocleavable fluorescent nucleotide analogues have been synthesized and characterized and have been used to produce 4-color DNA sequencing data on a chip. These nucleotides have been shown to be excellent substrates for the DNA polymerase and the fluorophore could be cleaved efficiently using near UV irradiation. This is important with respect to enhancing the speed of each cycle in SBS for high throughput DNA analysis. Also, a PCR-amplified DNA template can be ligated with a self-priming moiety demonstrated that and its sequence can be accurately identified in a DNA polymerase reaction on a chip, indicating that a PCR product from any organism can be potentially used as a template for the SBS system in the future. The modification of the 3'-OH of the photocleavable fluorescent nucleotide with a small chemical group to allow reversible termination may be considered. The library of photocleavable fluorescent nucleotides reported here will also facilitate single molecule DNA sequencing approaches.

Materials and Methods

All chemicals were purchased from Sigma-Aldrich unless otherwise indicated. ^1H NMR spectra were recorded on a Bruker 400 spectrometer. High-resolution MS (HRMS) data were obtained by using a JEOL JMS HX 110A mass spectrometer. Mass measurement of DNA was made on a Voyager DE matrix-assisted laser desorption ionization-time-of-flight (MALDI-TOF) mass spectrometer (Applied Biosystems). Photolysis was performed using a Spectra Physics GCR-150-30 Nd-YAG laser that generates light pulses at 355 nm (ca. 50 mJ/pulse, pulse length ca. 7 ns) at a frequency of 30 Hz with a light intensity at ca. 1.5 W/cm^2. The scanned fluorescence emission images were obtained by using a ScanArray Express scanner (Perkin-Elmer Life Sciences) equipped with four lasers with excitation wavelengths of 488, 543, 594, and 633 nm and emission filters centered at 522, 570, 614, and 670 nm.

Synthesis of photocleavable fluorescent nucleotides

Photocleavable fluorescent nucleotides dGTP-PC-Bodipy-FL-510, dUTP-PC-R6G, dATP-PC-ROX and dCTP-PC-Bodipy-650 (Fig. 6) were synthesized according to Figure 7 using a similar method as reported (16). A photocleavable linker (**PC-Linker**) 1-[5-(aminomethyl)-2-nitrophenyl]ethanol was reacted with the NHS ester of the corresponding fluorescent dye to produce an intermediate **PC-Dye**, which was converted to a **PC-Dye NHS ester** by reacting with *N, N'*-disuccinimidyl carbonate. The coupling reaction between the different **PC-Dye NHS esters** and the amino nucleotides (dATP-NH_2 and dGTP-NH_2 from Perkin-Elmer;

dUTP-NH_2 from Sigma; dCTP-NH_2 from TriLink BioTechnologies) produced the 4 photocleavable fluorescent nucleotides.

The synthesis of the photocleavable fluorescent nucleotide dCTP-PC-Bodipy-650 was reported (Seo, T. S., Bai, X., Ruparel, H., Li, Z., Turro, N. J. & Ju, J. (2004) *Proc. Natl. Acad. Sci. USA* **101**, 5488 5493). dUTP-PC-R6G, dGTP-PC-Bodipy-FL-510 and dATP-PC-ROX were prepared with a similar procedure as shown in Figure 7.

A general procedure for the synthesis of PC - Dye: 1-[5-(Aminomethyl)-2-nitrophenyl]ethanol (2) (PC-Linker, 5 mg, 26 μmol) was dissolved in 550 μl of acetonitrile (for Bodipy) or DMF (for R6G and ROX) and then mixed with 100 μl of 1 M $NaHCO_3$ aqueous solution. A solution of the NHS ester of the corresponding fluorophore (Molecular Probes) (13 μmol) in 400 μl of acetonitrile or DMF was added slowly to the above reaction mixture and then stirred for 5 h at room temperature. The resulting reaction mixture was purified on a preparative silica-gel TLC plate ($CHCl_3$/CH_3OH, 4/1 for Bodipy, 1/1 for R6G and ROX) to yield pure **PC-Dye**. **PC-R6G:** (92% yield) $_1$NMR (400 z, CD_3 D) 8.17 (d, 2H), 7.87(d, 2H),7.76 (d, 1H), 7.45(d, 1H), 7.04(s, 2H), 6.89(s, 2H), 5.34 (q, 1H), 4.70(s, 2H), 3.52(q, 4H), 2.14 (s, 6H), 1.47(d, 3H), 1.35 (t, 6H). HRMS (FAB_+) *m/z*: calcd for $C_{36}H_{37}N_4O_7$ (M + H_+), 637.2622; found, 637.2643. **PC-ROX:** (90% yield) $_1$NMR (400 z, CD_3 D) 8.11(d, 2H), 7.88 (m, 2H), 7.69 (d, 1H), 7.45 (dd, 1H), 6.79(s, 1H), 5.36(q, 1H), 4.69 (s, 2H), 3.50 (m, 8H), 3.08(t, 4H), 2.73 (t, 4H), 2.11 (t, 4H), 1.95 (t, 4H), 1.47 (d, 3H). HRMS (FAB_+) *m/z*: calcd for $C_{42}H_{41}N_4O_7$ (M + H_+),

713.2975; found, 713.2985. **PC-Bodipy-FL-510** was reported (Li, Z., Bai, X., Ruparel, H., Kim, S., Turro, N. J. & Ju, J. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 414-419.)

A general procedure for the synthesis of **PC-Dye NHS ester:** *N,N'*-disuccinimidyl carbonate (4.27 mg, 17 μmol) and triethylamine (4.6 μl, 33 μmol) were added to a solution of **PC-Dye** (11 μmol) in 200 μl of dry acetonitrile or DMF. The reaction mixture was stirred under argon at room temperature for 6 h. The solvent was removed under vacuum, and the residue was immediately purified by flash column chromatography (CH_2Cl_2/CH_3OH, 4/1). **PC-R6G NHS ester:** (28% yield) $_1$H NMR (400 MHz, CD_3OD) 8.15 (s, 2H), 8.03 (d, 1H), 7.78 (m, 1H), 7.71 (d, 1H), 7.56 (dd, 2H), 7.02 (m, 2H), 6.86 (m, 2H), 6.30 (q, 1H), 4.78 (d, 1H), 4.67 (d, 1H), 3.51 (q, 4H), 2.67 (s, 4H), 2.12 (s, 6H), 1.73 (d, 3H), 1.35 (t, 6H). HRMS (FAB_+) *m/z*: calculated for $C_{41}H_{40}O_{11}N_5$ ($M+H_+$), 778.2724; found, 778.2731. **PC-ROX NHS ester:** (35% yield) $_1$H NMR3 (400 MHz, CD_3OD) 8.09 (m, 2H), 8.02 (d, 1H), 7.69-7.75 (m, 2H), 7.54 (dd, 2H), 6.77 (m,2H), 6.30 (q, 1H), 4.78 (d, 1H), 4.66 (d, 1H), 3.47-3.57 (m, 8H), 3.04-3.10 (m, 4H), 2.64-2.72 (m, 8H), 2.06-2.14 (m, 4H), 1.90-1.98 (m, 4H), 1.74 (d, 3H). HRMS (FAB_+) *m/z*: calcd for $C_{47}H_{44}O_{11}N_5$ ($M+H_+$), 854.3037; found, 854.3069. **PC-Bodipy-FL-510 NHS ester** was reported (Li, Z., Bai, X., Ruparel, H., Kim, S., Turro, N. J. & Ju, J. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 414-419.)

A general procedure for the synthesis of photocleavable fluorescent nucleotides dUTP-PCR6G, dGTP-PC-Bodipy-FL-510 and dATP-PC-ROX

PC-Dye NHS ester (30 μmol) in 300 μl of acetonitrile or DMF was added to a solution of amino nucleotide dNTP-NH_2 (1 μmol) in 300 μl of 0.1 M Na_2CO_3-$NaHCO_3$ buffer (pH 8.5-8.7). The reaction mixture was stirred at room temperature for 3 h. A preparative silica-gel TLC plate was used to separate the unreacted **PCDye NHS ester** from the fractions containing final photocleavable fluorescent nucleotides ($CHCl_3$/CH_3OH, 1/1). The product was concentrated further under vacuum and purified with reverse-phase HPLC on a 150 4.6-mm C18 column to obtain the pure product. Mobile phase: A, 8.6 mM triethylamine/100 mM hexafluoroisopropyl alcohol in water (pH 8.1); B, methanol. Elution was performed with 100% A isocratic over 10 min followed by a linear gradient of 0-50% B for 20 min and then 50% B isocratic over another 20 min.

DNA polymerase reaction using 4 photocleavable fluorescent nucleotide analogues in solution

The four nucleotide analogues, dGTP-PC-Bodipy-FL-510, dUTP-PC-R6G, dATP-PC-ROX and dCTP-PC-Bodipy-650 were characterized by performing four continuous DNA-extension reactions sequentially using a primer (5'-AGAGGATCCAACCGAGAC-3') (SEQ ID NO:5) and a synthetic DNA template(5'-GTGTACATCAACATCACCTACCACCATGTCAGTCTCGGTTGGAT-CCTCTATTGTGTCCGG-3') (SEQ ID NO:6) corresponding to a portion of exon 7 of the human *p53* gene (Fig. 1). The

four nucleotides in the template immediately adjacent to the annealing site of the primer were 3′-ACTG-5′. First, a polymerase extension reaction using dUTP-PC-R6G along with the primer and the template was performed producing a single base extension product. The reaction mixture for this, and all subsequent extension reactions, consisted of 80 pmol of template, 50 pmol of primer, 80 pmol of the particular photocleavable fluorescent nucleotide, 1X Thermo Sequenase reaction buffer, and 4 U of Thermo Sequenase DNA polymerase (Amersham Biosciences) in a total volume of 20 μL. The reaction consisted of 25 cycles at 94 °C for 20 sec, 48 °C for 40 sec, and 60 °C for 75 sec. Subsequently, the extension product was purified by using reverse-phase HPLC. An Xterra MS C18 (4.6 x 50-mm) column (Waters) was used for the HPLC purification. Elution was performed over 120 minutes at a flow rate of 0.5 mL/min with the temperature set at 50 °C by using a linear gradient (12-34.5%) of methanol in a buffer consisting of 8.6 mM triethylamine and 100 mM hexafluoroisopropyl alcohol (pH 8.1). The fraction containing the desired DNA product was collected and freeze-dried for analysis using MALDI-TOF mass spectrometry. For photocleavage, the purified DNA extension product bearing the fluorescent nucleotide analogue was resuspended in 200 μL of deionized water. The mixture was irradiated for 10 seconds in a quartz cell with path lengths of 1.0 cm employing a Nd-YAG laser at 355 nm and then analyzed by MALDI-TOF MS. After photocleavage, the DNA product with the fluorophore removed was used as a primer for a second extension reaction using dGTP-PC-Bodipy-FL-510. The second extended

product was then purified by HPLC and photolyzed. The third extension using dATP-PC-ROX and the fourth extension using dCTP-PC-Bodipy-650 were carried out in a similar manner using the previously extended and photocleaved product as the primer.

PCR amplification to produce azido-labeled DNA template

An azido-labeled PCR product was obtained using a 100-bp template (5'-AGCGACTGCTATCATGTCATATCGACGTGCTCACTAGCTCTAC ATATGCGTGCGTGATCAGATGACGTATCGATGCTGACTATAGTCTCCCATGCGAGTG -3'), (SEQ ID NO:7) a 24-bp azido-labeled forward primer (5'-N_3-AGCGACTGCTATCATGTCATATCG-3'), (SEQ ID NO:8) and a 24-bp unlabeled reverse primer (5'-CACTCGCATGGGAGACTATAGTCA-3'). (SEQ ID NO:9) In a total reaction volume of 50 μL, 1 pmol of template and 30 pmol of forward and reverse primers were mixed with 1 U of AccuPrime *Pfx* DNA polymerase and 5 μL of 10X AccuPrime *Pfx* reaction mix (Invitrogen) containing 1 mM of $MgSO_4$ and 0.3 mM of dNTP. The PCR reaction consisted of an initial denaturation step at 95 °C for 1 min, followed by 38 cycles at 94 °C for 15 sec, 63 °C for 30 sec, 68 °C for 30 sec. The product was purified using a 96 QlAquick multiwell PCR purification kit (Qiagen) and the quality was checked using 2% agarose gel electrophoresis in 1X TAE buffer. The concentration of the purified PCR product was measured using a Perkin-Elmer Lambda 40 UV-Vis spectrophotometer.

Construction of a self-priming DNA template on a chip by enzymatic ligation

The amino-modified glass slide (Sigma) was functionalized to contain a terminal alkynyl group as described previously (16). The azido-labeled DNA product generated by PCR was dissolved in DMSO/H_2O (1/3, v/v) to obtain a 20 μM solution. 5 μL of the DNA solution was mixed with CuI (10 nmol, 100 eq.) and *N,N*-diisopropyl-ethylamine (DIPEA) (10 nmol, 100 eq.) and then spotted onto the alkynyl-modified glass surface in the form of 6 μL drops. The glass slide was incubated in a humid chamber at room temperature for 24 hr, washed with deionized water (dH_2O) and SPSC buffer (50 mM sodium phosphate, 1 M NaCl, pH 6.5) for 1 hr (16), and finally rinsed with dH_2O. To denature the double stranded PCR-amplified DNA to remove the non-azido-labeled strand, the glass slide was immersed into 0.1 M NaOH solution for 10 min and then washed with 0.1 M NaOH and dH_2O, producing a single stranded DNA template that is immobilized on the chip. For the enzymatic ligation of a self-priming moiety to the immobilized DNA template on the chip, a 5'-phosphorylated 40-bp loop primer (5'-PO_3-GCTGAATTCCGCGTTCGCGGAATTCAGCCACTCGCATGGG-3') (SEQ ID NO:10) was synthesized. This primer contained a thermally stable loop sequence 3'-G(CTTG)C-5', a 12-bp stem, and a 12-bp overhanging end that would be annealed to the immobilized single stranded template at its 3'-end. A 10 μL solution consisting of 100 pmol of the primer, 10 U of *Taq* DNA ligase, 0.1 mM NAD, and 1X reaction buffer (New England Biolabs) was spotted onto a location of the chip

containing the immobilized DNA and incubated at 45 °C for 4 hr. The glass slide was washed with dH_2O, SPSC buffer, and again with dH_2O. The formation of a stable hairpin was ascertained by covering the entire surface with 1X reaction buffer (26 mM Tris·HCl/6.5 mM $MgCl_2$, pH 9.3), incubating it in a humid chamber at 94 °C for 5 min to dissociate any partial hairpin structure, and then slowly cooling down to room temperature for reannealing.

SBS reaction on a chip with four photocleavable fluorescent nucleotide analogues

One microliter of a solution consisting of dATP-PC-ROX (60 pmol), 2 U of Thermo Sequenase DNA polymerase, and 1X reaction buffer was spotted on the surface of the chip, where the self-primed DNA moiety was immobilized. The nucleotide analogue was allowed to incorporate into the primer at 72 °C for 5 min. After washing with a mixture of SPSC buffer, 0.1% SDS, and 0.1% Tween 20 for 10 min, the surface was rinsed with dH_2O and ethanol successively, and then scanned with a ScanArray Express scanner to detect the fluorescence signal. To perform photocleavage, the glass chip was placed inside a chamber (50 x 50 x 50 mm) filled with acetonitrile/water (1/1, v/v) solution and irradiated for 1 min with the Nd-YAG laser at 355 nm. The light intensity applied on the glass surface was ca. 1.5 W/cm^2. After washing the surface with dH_2O and ethanol, the surface was scanned again to compare the intensity of fluorescence after photocleavage with the original fluorescence intensity. This process was followed by the incorporation of dGTP-PC-Bodipy-FL-510, with the subsequent washing, fluorescence detection,

and photocleavage processes performed as described above. The same cycle was repeated 10 more times using each of the four photocleavable fluorescent nucleotide analogues complementary to the base on the template. For a negative control experiment, 1 μL solution containing dATP-PC-ROX (60 pmol), and 1X reaction buffer was added on to the DNA immobilized on the chip in the absence of DNA polymerase and then incubated at 72 °C for 5 min, followed by the same washing and detection steps as above.

References

1. Collins, F. S., Green, E. D., Guttmacher, A. E. & Guyer, M. S. (2003) *Nature* **422**, 835-847.
2. Thomas, J. W., Touchman, J. W., Blakesley, R. W., Bouffard, G. G., Beckstrom-Sternberg, S. M., Margulies, E. H., Blanchette, M., Siepel, A. C., Thomas, P. J. & McDowell, J. C. *et al.* (2003) *Nature* **424**, 788-793.
3. Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B. H. & Hood, L. E. (1987) *Nature* **321**, 674-679.
4. Ju, J., Ruan, C., Fuller, C. W., Glazer, A. N. & Mathies, R. A. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 4347-4351.
5. Doherty, E.A.S., Kan, C. W. and Barron, A. E. (2003) *Electrophoresis*, **24**, 4170-4180.
6. Drmanac, S., Kita, D., Labat, I., Hauser, B., Schmidt, C., Burczak, J. D. & Drmanac, R. (1998) *Nat. Biotechnol.* **16**, 54-58.
7. Fu, D. J., Tang, K., Braun, A., Reuter, D., Darnhofer-Demar, B., Little, D. P., O'Donnell, M. J., Cantor, C. R. & Koster, H. (1998) *Nat. Biotechnol.* **16**, 381-384.
8. Roskey, M. T., Juhasz, P., Smirnov, I. P., Takach, E. J., Martin, S. A. & Haff, L. A. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 4724-4729.
9. Edwards, J. R., Itagaki, Y. & Ju, J. (2001) *Nucleic Acids Res.* **29**, e104 (p1-6).

10. Kasianowicz, J. J., Brandin, E., Branton, D. & Deamer, D. W. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 13770–13773.

11. Ronaghi, M., Uhlen, M. & Nyren, P. (1998) *Science* **281**, 363–365.

12. Braslavsky, I., Hebert, B., Kartalov, E. & Quake, S. R. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 3960–3964.

13. Mitra, R. D., Shendure, J., Olejnik, J., Olejnik, E. K. & Church, G. M. (2003) *Anal. Biochem.* **320**, 55–65.

14. Hyman, E. D. (1988) *Anal. Biochem.* **174**, 423–436.

15. Ju, J., Li, Z., Edwards, J. & Itagaki, Y. (2003) U. S. Patent 6,664,079.

16. Seo, T. S., Bai, X., Ruparel, H., Li, Z., Turro, N. J. & Ju, J. (2004) *Proc. Natl. Acad. Sci. USA*, **101**, 5488–5493.

17. Rosenblum, B. B., Lee, L. G., Spurgeon, S. L., Khan, S. H., Menchen, S. M., Heiner, C. R. & Chen, S. M. (1997) *Nucleic Acids Res.* **25**, 4500–4504.

18. Zhu, Z., Chao, J., Yu, H. & Waggoner, A. S. (1994) *Nucleic Acids Res.* **22**, 3418–3422

19. Rajasekharan Pillai, V. N. (1980) *Synthesis* **1**, 1–26.

20. Barany, F. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 189–193.

21. Antao, V. P., Lai, S. Y. & Tinoco, I. Jr. (1991) *Nucleic Acids Res.* **19**, 5901–5905.